Adaptability of machine learning methods and hydrological models to discharge simulations in data-sparse glaciated watersheds

JI Huiping[1,2], CHEN Yaning[1,2*], FANG Gonghuan[1,2], LI Zhi[1,2], DUAN Weili[1,2], ZHANG Qifei[1,2]

[1] State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China;
[2] University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: The accurate simulation and prediction of runoff in alpine glaciated watersheds is of increasing importance for the comprehensive management and utilization of water resources. In this study, long short-term memory (LSTM), a state-of-the-art artificial neural network algorithm, is applied to simulate the daily discharge of two data-sparse glaciated watersheds in the Tianshan Mountains in Central Asia. Two other classic machine learning methods, namely extreme gradient boosting (XGBoost) and support vector regression (SVR), along with a distributed hydrological model (Soil and Water Assessment Tool (SWAT) and an extended SWAT model (SWAT_Glacier) are also employed for comparison. This paper aims to provide an efficient and reliable method for simulating discharge in glaciated alpine regions that have insufficient observed meteorological data. The two typical basins in this study are the main tributaries (the Kumaric and Toxkan rivers) of the Aksu River in the south Tianshan Mountains, which are dominated by snow and glacier meltwater and precipitation. Our comparative analysis indicates that simulations from the LSTM shows the best agreement with the observations. The performance metrics Nash-Sutcliffe efficiency coefficient (NS) and correlation coefficient (R^2) of LSTM are higher than 0.90 in both the training and testing periods in the Kumaric River Basin, and NS and R^2 are also higher than 0.70 in the Toxkan River Basin. Compared to classic machine learning algorithms, LSTM shows significant advantages over most evaluating indices. XGBoost also has high NS value in the training period, but is prone to overfitting the discharge. Compared with the widely used hydrological models, LSTM has advantages in predicting accuracy, despite having fewer data inputs. Moreover, LSTM only requires meteorological data rather than physical characteristics of underlying data. As an extension of SWAT, the SWAT_Glacier model shows good adaptability in discharge simulation, outperforming the original SWAT model, but at the cost of increasing the complexity of the model. Compared with the oftentimes complex semi-distributed physical hydrological models, the LSTM method not only eliminates the tedious calibration process of hydrological parameters, but also significantly reduces the calculation time and costs. Overall, LSTM shows immense promise in dealing with scarce meteorological data in glaciated catchments.

Keywords: hydrological simulation; long short-term memory; extreme gradient boosting; support vector regression; SWAT_Glacier model; Tianshan Mountains

*Corresponding author: CHEN Yaning (E-mail: chenyn@ms.xjb.ac.cn)
Received 2021-04-07; revised 2021-04-29; accepted 2021-05-08

1 Introduction

The Tianshan Mountains are considered as the "water tower of Central Asia". The mountains serve as the source for numerous rivers in the five Central Asian countries and Northwest China, providing freshwater for hundreds of millions of people in the lower reaches of the rivers. In so doing, the Tianshan Mountains ensure food security, economic development, water consumption, and ecological water utilization for the entire region, while simultaneously playing an important role as an ecological barrier (Chen et al., 2016; Immerzeel et al., 2020; Viviroli et al., 2020). However, with the hard-to-reach terrain and harsh weather conditions, the availability of observed data is scarce in the Central Asian alpine areas. Hydrological simulation is thus immensely challenging but critically important under increasingly worsening global warming conditions (Immerzeel et al., 2010; Duethmann et al., 2016; Viviroli et al., 2020).

Hydrologists have committed to developing various hydrological simulation approaches, including applying traditional statistical methods based on mathematics or a variety of hydrological models for dynamic simulation (Chen et al., 2006; Finger et al., 2015; Shen et al., 2018a; Greve et al., 2020). Statistical methods are widely used to describe the static state of a natural system, while hydrological modeling is a dynamic description of complex hydrological processes (Chen et al., 2017; Shen et al., 2018b). Over the past several decades, hydrological modeling has experienced a long development process and is now broadly applied-used to explain various hydrological processes, along with future flood and runoff predictions (Tarasova et al., 2016; Jodar et al., 2018). Duethmann et al. (2015) investigated variations in runoff at the Sari-Djaz watershed by applying the Water Availability in Semi-Arid Environments (WASA) model; they attributed most of the increase in runoff to a surge in glacier melt. Chen et al. (2017) reviewed the status of hydrological modeling in glaciated catchments and discussed the limitations of the available models. Previous studies found that the performance of hydrological models is closely related to the spatial and temporal resolution of model-driven data. The quality of the meteorological inputs plays a leading role, such that even minor data errors or noise interference may lead to large deviations (Michaud and Sorooshian, 1994; Ajami et al., 2004). In this regard, the scarce and unevenly-distributed meteorological stations in the Tianshan Mountains makes it challenging to simulate runoff in the alpine watershed dominated by glacier and snow meltwater and precipitation (Qin et al., 2009; Chen et al., 2017). Therefore, it is difficult to improve the precision of discharge simulations solely through the use of hydrological models.

With the rapid development of artificial intelligence, researchers are exploring machine learning methods for discharge simulation (Maier et al., 2010; Fang et al., 2017; Shen et al., 2018; Fang and Shen, 2019; Yang et al., 2019). Compared with physical hydrological modeling, data-driven models ignore the physical logic relationship between variables and pay more attention to the potential relationship of the extremely large data sample for training in order to improve the prediction accuracy (Reichstein et al., 2019; Rudin, 2019). Artificial neural networks (ANNs), which were first developed by Hsu et al. (1995), provide significant advantages in hydrological simulations. Long short-term memory (LSTM), which is a machine learning algorithm for time series modeling, has also recently emerged as a powerful tool for capturing variations in discharge (Shen, 2018; Feng et al., 2020; Fu et al., 2020; Hu et al., 2020; Li et al., 2021). Hu et al. (2018) discovered that the performance of LSTM is significantly better than that of the artificial neural network (ANN) model for flood predictions in the Fenhe River Basin, China. Gao et al. (2020) used LSTM to perform flood predictions for the Shaxi River Basin in China, finding that LSTM has a high prediction accuracy at different time steps. Xiang et al. (2020) developed a LSTM sequence-to-sequence model to estimate hourly rainfall-runoff in two midwestern watersheds in Iowa, USA. Meanwhile, Kratzert et al. (2020) used different precipitation products as inputs to the LSTM model to improve discharge simulations, revealing that the model is able to extract information from the characteristics of watersheds so as to distinguish rainfall-runoff in different watersheds; they further applied this concept to runoff predictions in unmeasured watersheds. These results showed that LSTM has unprecedented accuracy in discharge simulations and predictions in ungauged basins (Kratzert et al., 2018, 2019a, b).

In this paper, we explore the capability of LSTM in simulating discharge in the data-scarce Kumaric and Toxkan rivers of the Aksu River Basin in the Tianshan Mountains, and compare it with the distributed hydrological models and machine learning methods, i.e., extreme gradient boosting (XGBoost) and support vector regression (SVR). The aim is to compare the results of the three machine learning methods (LSTM, XGBoost, and SVR) with the classical semi-distributed hydrological models (Soil and Water Assessment Tool (SWAT) and SWAT_Glacier (an extended SWAT model with glacier melting mechanism)) to determine an efficient method for discharge forecasting in glaciated alpine catchments with complex hydrological mechanisms.

2 Study area and data sources

2.1 Study area

The Aksu River originates in the south Tianshan Mountains and provides more than 70% of the water for the mainstream of the Tarim River (Fan et al., 2014). The Aksu River Basin is not only responsible for providing living and irrigation water, but also for the maintenance of the "green corridor" in the lower reaches of the Tarim River. The total area of the Aksu River Basin is 6.30×10^4 km^2, and the plain area covers 2.90×10^4 km^2. Recent data indicate that there are 2740 glaciers of various sizes in the alpine regions, covering an area of 0.49×10^4 km^2. The Kumaric and Toxkan rivers converge in Wensu County to form the Aksu River, giving the river a total length of 224 km.

The Kumaric River Basin with the outlet hydrological station of Xiehela, has an area of 1.29×10^4 km^2 and ranges in altitude from 1435 to 7126 m. There are three meteorological stations in the watershed, namely Aksu, Koilu, and Tianshan. The hydrological station for the Toxkan River Basin is called Shaliguilanke, and the two meteorological stations near the river are Akqi and Tuergate, as shown in Figure 1.

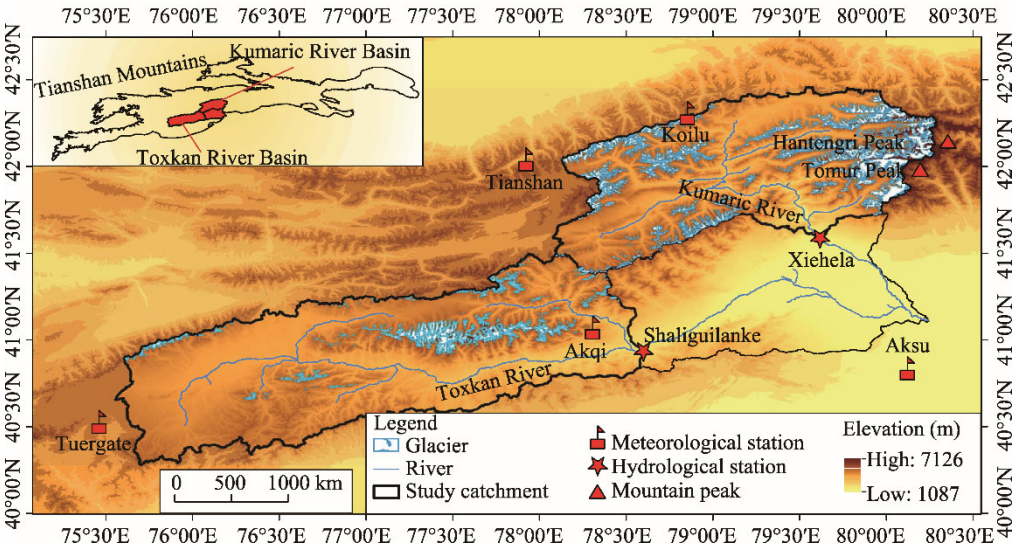

Fig. 1 Location of Kumaric and Toxkan river basins as well as distributions of hydrological stations, meteorological stations, and glacier distribution in the Tianshan Mountains

2.2 Datasets

Daily meteorological data include maximum temperature (°C), minimum temperature (°C), precipitation (mm), solar radiation (KJ/(m^2·d), average wind speed (m/s), and relative humidity (%). The data of meteorological stations which located in China are from the National Meteorological Information Center (http://data.cma.cn/), while the data of the Koilu and Tianshan stations are interpolated on the basis of two datasets: APHRODITE (Asian Precipitation Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources) and CRU

(Climatic Research Unit). All the meteorological data are from 2002 to 2011. Compared to various meteorological datasets with different time scales synthesized by a variety of spatial interpolation or assimilation methods, the measured data from meteorological observation stations are considered the most reliable, especially in the complex terrain of the Tianshan Mountains. For this reason, the data obtained from observation stations in the basins under study are selected as driving data for the machine learning methods and hydrological models.

The observed discharge data from 2002 to 2011 at two of the hydrological stations were obtained from the Tarim River Basin Management Bureau of Xinjiang Uygur Autonomous Region, China. Along with meteorological data, the spatial data, including digital elevation model (DEM) data, land use, soil data, and glacier and snow distribution data as described in Ji et al. (2019), are used as inputs in the two hydrological models.

3 Methodology

3.1 Long short-term memory (LSTM)

The LSTM model is an optimization improvement developed by Hochreiter and Schmidhuber (1997), based on a traditional recurrent neural network (RNN). The application of RNN was previously limited in long-time series data by its finite memory ability. When the error "backward propagates" from the last timestamp to the beginning timestamp, the gradient decays or grows exponentially. Moreover, if the time series of the training data is too long, the propagated errors from the last timestamp either reduce to 0 or expand to NaN when they arrive at the beginning timestamp. Also, the model cannot converge, which leads to training failure (Greff et al., 2017). The discharge under study shows seasonal variations within a year and also has a lag effect of several days in response to precipitation. This represents a significant challenge or even an impossibility for RNN, which has only ten days of very short memory ability to simulate discharge (Bengio et al., 1994). LSTM, as a representative deep learning model, effectively solves the problem of RNN being unable to remember long-time series.

Compared with standard RNN, a self-evaluation mechanism is a unique feature of LSTM. By weighing and comparing its own value and the importance of the information in memory, it chooses to discard unimportant information, which is controlled by the added gates in each cell. The internal structure of each cell is relatively complex and includes the input gate, forget gate, cell state, and output gate, as shown in Figure 2.

3.1.1 Forget gate layer

The forget gate plays a very important role in LSTM model training. In fact, it is the essence of LSTM design. LSTM can selectively memorize important information, which is mainly realized by the forget gate. The hidden layer output result H_{t-1} of the previous time t–1 is transferred to the cell at time t, and the input value X_t at time t enters the forget gate along with it. The responsibility of the forget gate is to control the extent to which H_{t-1} and X_t are forgotten; the calculation method is as follows:

$$F_t = \sigma\left(W_f \times \left[H_{t-1}, X_t + b_f\right]\right), \tag{1}$$

where F_t is the value of forget gate at time step t, and the range of F_t is 0–1; σ is the sigmoid activation function; W_f and b_f represent the weight and deviation of the forget gate, respectively; H_{t-1} is the hidden layer output result of the previous time t–1; and X_t is the input value at time t.

Furthermore, the forget gate measures its value by each input. If the input is not beneficial to the whole sequence information, it will be discarded as unimportant. If the value of F_t is close to 1, the input value X_t at time t will be forgotten, and vice versa, in which case F_t will be calculated and passed to the cell state. LSTM uses selective forgetting to store effective information to ensure long-term memory, rather than the total absorption nature of the RNN method.

3.1.2 Input gate layer

The input gate controls which important information will be updated and stored in the memory unit. The calculation formulas are shown in Equations 2 and 3.

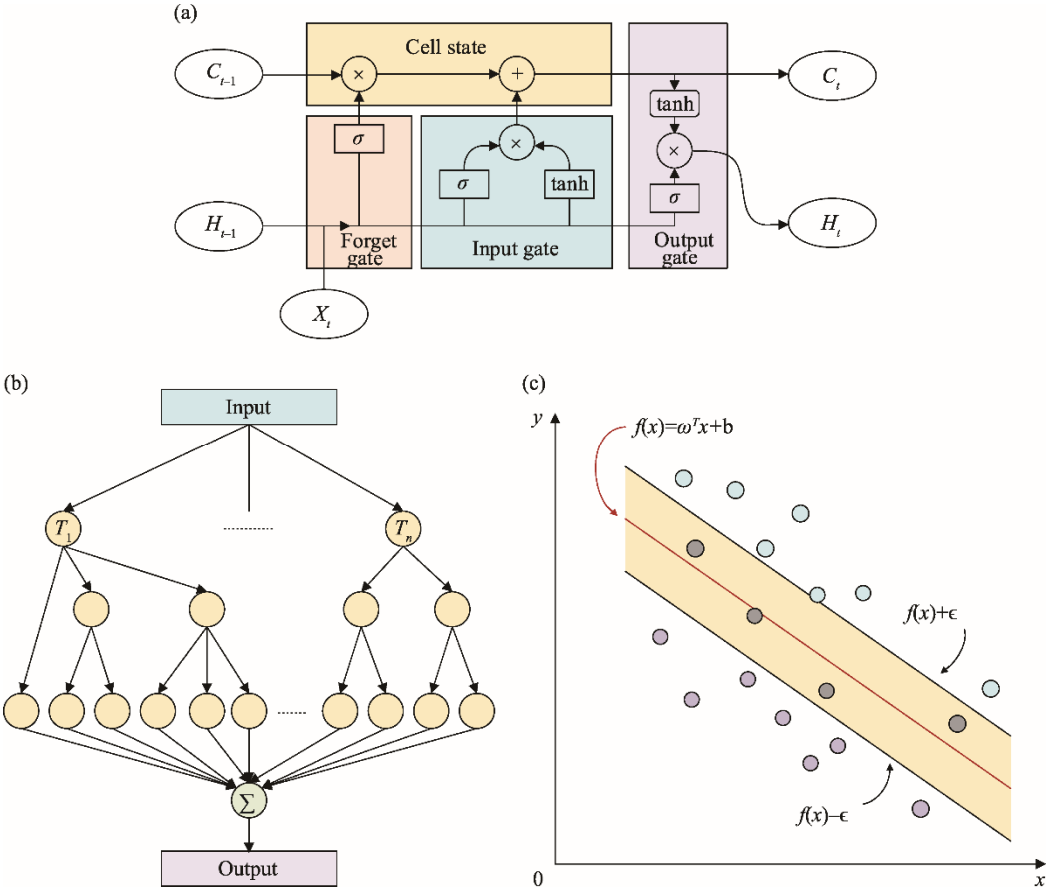

Fig. 2 Basic structure of long short-term memory (LSTM) and detailed calculation process at time step t (a), structural diagram of extreme gradient boosting (XGBoost; b), and diagram of support vector regression (SVR; c). The yellow highlighted area indicates the interval band of 2ϵ, where ϵ means the absolute deviation value. C_t, value of cell state at time step t; C_{t-1}, value of cell state at time step t–1; X_t, current input; H_t, output result of hidden layer at time step t; H_{t-1}, hidden layer output result of the previous time t–1; tanh, hyperbolic tangent function; σ, sigmoid activation function; T_1, the 1st tree; T_n, the n^{th} tree; $f(x)$, the function of the model.

$$I_t = \sigma\left(W_i \times \left[H_{t-1}, X_t + b_i\right]\right), \tag{2}$$

$$\widetilde{C}_t = \tanh\left(W_c \times \left[H_{t-1}, X_t\right] + b_c\right), \tag{3}$$

where I_t is the value of input gate at time step t, and it is also calculated by the activation function sigmoid, with a value range of 0–1; W_i and b_i represent the weight and deviation of the input gate, respectively; $\widetilde{C}_t$ is the cell update candidate; tanh means the hyperbolic tangent function; and W_c and b_c represent the weight and deviations of the cell, respectively. The purpose of introducing the cell update candidate $\widetilde{C}_t$ is to multiply its calculation results by I_t and pass to the cell state as the output of the input gate.

3.1.3 Cell state

The cell state is used in combination with the forget gate layer to store information. By comparing the value of the input result at time t and the output result of the forget gate, it can be determined whether to pass more previous output results to the output gate or add more current input information to the sequence. This step is to ensure that important information is saved and redundant or unimportant information is discarded. Updating calculation of C_t is shown in Equation 4:

$$C_t = F_t \times C_{t-1} + I_t \times \widetilde{C}_t, \tag{4}$$

where C_t is the value of cell state at time step t; C_{t-1} is the value of cell state of the previous time t–1; and the updated C_t at time t+1 will be passed to the next cell as the input.

3.1.4 Output gate layer

The output gate controls how the information flow of the cell state at time t enters the hidden layer at time t+1. The output gate also uses its defined weight W_o and deviation b_o to calculate O_t. The calculation formulas are as follows:

$$O_t = \sigma\left(W_o \times \left[H_{t-1}, X_t + b_o\right]\right), \tag{5}$$

$$H_t = O_t \times \tanh\left(C_t\right), \tag{6}$$

where O_t is the value of output gate at time step t; W_o and b_o represent the weight and deviation of the output gate, respectively; and H_t is the output result of the hidden layer at time t.

Each gate of LSTM has its own unique parameter group weight term W and bias term b. Additionally, there is a weighted sum of the new input and the output of the previous hidden layer in each cell, which shows significant long-term memory characteristics and is easier to train. The most important facet is to make up for the shortcomings of the traditional RNN model. The activation function endows neural networks with nonlinear characteristics. Therefore, the correct selection of activation function can improve the data processing ability of these networks and make the simulation results closer to the observation values. It can also improve the computational efficiency of the networks. The initialization of weights is affected by random starting values, which leads to the uncertainty of LSTM in training. In this study, we selected Adam as the stochastic optimization method (Kingma and Ba, 2015).

3.2 Extreme gradient boosting (XGBoost)

XGBoost is a new scalable machine learning algorithm developed by Chen and Guestrin (2016). Gradient Boosting Decision Tree (GBDT) is an iterative decision tree algorithm that uses the gradient descent method to minimize the loss function in the parameter space to solve the optimization problem, and XGBoost is a modified form of it. XGBoost is widely used mainly because of its regularized boosting technology. The regularization term is a significant characteristic of XGBoost, which is superior to traditional GBDT and can be considered a penalty for the complexity level of the model. By evaluating the complexity feature of each regression tree, the complexity of the model can be effectively controlled to prevent overfitting. The objective function Obj is composed of error function and penalty, as shown in the following formulas:

$$Obj = \sum_{i=1}^{n} l\left(y_i - \hat{y}_i\right) + \sum_{t=1}^{t} \Omega\left(f_t\right), \tag{7}$$

$$\Omega\left(f_t\right) = \gamma T + \frac{1}{2}\lambda\left\|\omega\right\|^2, \tag{8}$$

where Obj is the objective function; l is the loss function between the observed value y_i and the predicted value $\hat{y}_i$ of the model; Ω is the penalty; γ and λ are two adjustable parameters; T represents the number of leaf nodes; and ω is the leaf node fraction.

XGBoost chooses Newton's method to solve the optimization problem in the function space and can customize the objective function according to the user's needs. Calculating the error function requires second-order Taylor expansion (Zheng et al., 2019; Ni et al., 2020). Therefore, the selection of the objective function needs to meet the second-order differentiable condition. XGBoost is essentially an additive training model, and each regression tree learns the results and errors of all previous tree iterations. It is also a training process based on residuals. Because of its good performance in computing speed and simulation precision, XGBoost is widely used in hydrology research and is one of the most popular machine learning models. In this study, we applied XGBoost to simulate daily discharge in glaciated alpine watersheds in order to test its performance.

3.3 Support vector regression (SVR)

SVR is the regression branch of support vector machine (Yang et al., 2017). As a classical

regression model, SVR is a kernel-based algorithm. It can be employed to realize the nonlinearity of decision function and effectively fits the nonlinear relationship of samples (Aggarwal et al., 2012; Cheng et al., 2015; Luo et al., 2019). The difference between SVR and general linear regression is that it designs a spacer band on both sides of the linear function to allow some deviation between the model output value and the observed value. If the absolute deviation value is less than ϵ (absolute deviation value), the data located in the spacer band and deviation will be ignored, whereas if the absolute value of the deviation is greater than ϵ, the loss needs to be calculated. Therefore, the spacer band of 2ϵ is considered a loss-free area, as shown in Figure 2c.

In addition, while traditional linear regression algorithms optimize a model by calculating the mean value after gradient descent, SVR obtains the optimized model by maximizing the width of the spacer band and minimizing the total loss. The penalty factor C is an important parameter for the SVR algorithm to control overfitting. The value of C represents the importance of outliers and can also be interpreted as the tolerance of errors. An increase in the C value represents a smaller range of allowable error and a greater possibility of overfitting. Conversely, if the C value is close to 0, it is prone to underfitting. The C value essentially controls the generalization ability of the model. If the C value is too small or too large, the generalization ability of the model worsens.

Within the setting of the SVR algorithm, the coefficient gamma is an important parameter of the Radial Basis Function Kernel, which determines the distribution of model input data mapped to a new high-dimensional feature space. The gamma value affects the running speed of the model and must be greater than 0. The larger the gamma value, the fewer the number of support vectors. However, if the gamma value is too large, it is easy to generalize errors, leading to the phenomenon of overfitting.

3.4 Hydrological models

The SWAT is a classic semi-distributed hydrological model (Arnold et al., 1998) that is well-suited for accurately describing the hydrological process in catchments by transferring information between hydrological response units (Li et al., 2010). The SWAT model with a snow-melting module has a good applicability in snow-covered areas (Chen et al., 2017), but the model's suitability is limited in high-altitude catchments where glaciers are widely distributed. In these cases, the degree-day method is employed to construct a glacier-melting module and integrate it into the SWAT model. The hydrological model with a melting mechanism is called the SWAT_Glacier model. It has been applied in recent studies to simulate the glacier melting processes in glaciated watersheds in Central Asia. Details on the SWAT and SWAT_Glacier models can be found in Ji et al. (2019).

The three above-mentioned machine learning methods (LSTM, XGBoost, and SVR) and two hydrological models (SWAT and SWAT_Glacier) based on physical mechanisms are popular in hydrological research. In this study, we attempt to systematically apply them to simulate the daily discharge in data-sparse glaciated river basins. The input data of machine learning algorithms include daily maximum and minimum temperatures and precipitation. The observed discharge data are used to train and validate the models. To compare the performance of the five models, the time-setting of the training and testing periods for the data-driven models is consistent with the two hydrological models. The training period is from 1 January 2002 to 31 December 2007, and the testing period is from 1 January 2008 to 31 December 2011.

3.5 Performance metrics

We selected four measurement indices, i.e., Nash-Sutcliffe efficiency coefficient (NS), percentage bias (PBIAS), correlation coefficient (R^2), and root mean square error (RMSE), to evaluate and compare the capability of machine learning algorithms and hydrological models in simulating hydrological processes.

$$\mathrm{NS} = 1 - \frac{\sum_{i=1}^{n}\left(Q_{\mathrm{obs}}^{i} - Q_{\mathrm{sim}}^{i}\right)^2}{\sum_{i=1}^{n}\left(Q_{\mathrm{obs}}^{i} - \overline{Q_{\mathrm{obs}}}\right)^2}, \tag{9}$$

$$R^2 = \left\{ \frac{\sum_{i=1}^{n}\left(Q_{\text{obs}}^{i} - \overline{Q_{\text{obs}}}\right)\times\left(Q_{\text{sim}}^{i} - \overline{Q_{\text{sim}}}\right)}{\sqrt{\sum_{i=1}^{n}\left(Q_{\text{obs}}^{i} - \overline{Q_{\text{obs}}}\right)^2 \times \left[\sum_{i=1}^{n}\left(Q_{\text{sim}}^{i} - \overline{Q_{\text{sim}}}\right)^2\right]}} \right\}^2, \tag{10}$$

$$\text{PBIAS} = \frac{\sum_{i=1}^{n}\left(Q_{\text{sim}}^{i} - Q_{\text{obs}}^{i}\right)}{\sum_{i=1}^{n}\left(Q_{\text{obs}}^{i}\right)}, \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(Q_{\text{obs}}^{i} - Q_{\text{sim}}^{i}\right)^2}{n}}, \tag{12}$$

where n is the length of observation or simulation; Q_{obs}^{i} and Q_{sim}^{i} are the observed and simulated discharges at time step i (m^3/s); and $\overline{Q_{\text{obs}}}$ and $\overline{Q_{\text{sim}}}$ are the averages of the observed and simulated discharges (m^3/s), respectively.

For these four measures, NS ranges from negative infinitesimal to 1.00, with NS=1.00 meaning perfect simulation. NS is largely influenced by peak flows. R^2 represents the collinearity of observed and simulated data, and PBIAS appraises the average deviation of the simulation from the observation, with negative values indicating an underestimation and vice versa. A value of zero for PBIAS is ideal, because it means there is no deviation. A small RMSE value reveals the good fitting performance and precise predictability of the model.

4 Results

In this study, we drew hydrographs of the observed and simulated discharges of the Kumaric and Toxkan river basins to compare the performances of the three machine learning methods (LSTM, XGBoost, and SVR) and two hydrological models (SWAT and SWAT_Glacier). Scatterplots further illustrate the correlations between the simulations and observations. Figures 3 and 4 showed the simulation results of daily discharge in the Kumaric and Toxkan river basins, respectively.

As can be seen in Figures 3 and 4, all of the models obtained satisfactory results in the simulations of baseflow. However, their performance in the simulations of peak flow was quite different. The LSTM model did an excellent job of reproducing the observations, and its performance was significantly better than that of either XGBoost or SVR in terms of PBIAS (Table 1). In comparing the hydrographs of the two hydrological models, we can conclude that the SWAT_Glacier model had a better performance than the SWAT model, especially for the summer peak flow. Furthermore, the SWAT_Glacier model was able to effectively simulate daily discharge, while the hydrograph of the SWAT model was relatively flat. This indicated that it is difficult to capture changes in discharge caused by increases in glacier and snow meltwater over a short-time period due to temperature rises. This phenomenon was more pronounced in the validation period (Fig. 5). A comprehensive comparison showed that LSTM has significant advantages, demonstrating a good fit with the observations. In contrast, the SWAT_Glacier and SWAT models did not perform as well in the testing period as in the training period, and their fitting degree for the Toxkan River Basin significantly decreased with large deviations.

The hydrographs showed the performance of each model intuitively. However, for enhanced specificity, the statistical indices of the above five methods for both the training and testing periods are summarized for further comparison, as shown in Table 1.

For the Kumaric River Basin, a comparison of the evaluation indices of the five methods revealed that LSTM was significantly better than the others in terms of the correlation index and relative deviation index. Specifically, the performance of LSTM was satisfactory and had only a slightly underestimated peak discharge during the training period, with NS being 0.96 and PBIAS of –0.78%. The results of the hydrological models were not as good as LSTM in terms of evaluation indices. For the SWAT_Glacier model, the indices were slightly poorer than the LSTM model, with

NS=0.82, R^2=0.83, PBIAS=0.94%, and RMSE showing a value of 85.07 during the training period.

Figure 5 represented the predictions of the three machine learning algorithms and two hydrological models for discharge during the validation period. As can be seen from Figure 5a and b, LSTM performed excellently, with all performance indicators significantly better than those for

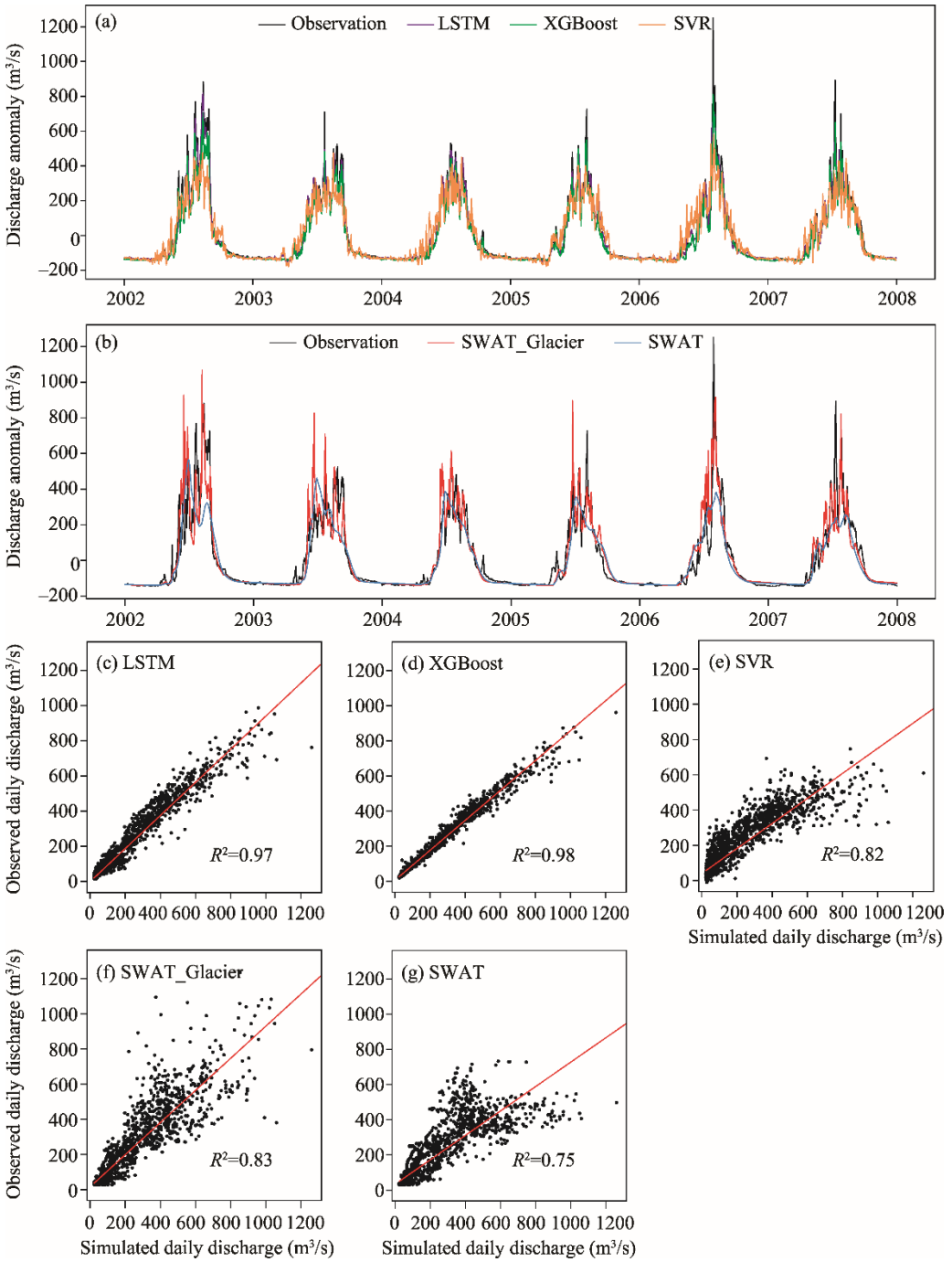

Fig. 3 Observed and simulated discharges of the five models for the Kumaric River Basin during the training period. (a), daily observations and simulations of the machine learning methods (LSTM, XGBoost, and SVR); (b), simulation performance of the traditional hydrological models (Soil and Water Assessment Tool (SWAT) and SWAT_Glacier (an extended SWAT model with glacier melting mechanism)); (c)–(g), scatterplots of observed and simulated values for each model, with R^2 representing the correlation coefficient.

the other models. For LSTM, NS and R^2 were greater than 0.90, which is very high for the prediction of daily discharge in an alpine basin. Furthermore, PBIAS was only –2.60%, RMSE was 60.26, and a relatively minor deviation was obtained. In contrast, for the SWAT_Glacier model, the indices were lower than those for the LSTM model, with NS=0.79, R^2=0.80, PBIAS= –5.4%, and RMSE showing a large value of 83.06 during the testing period.

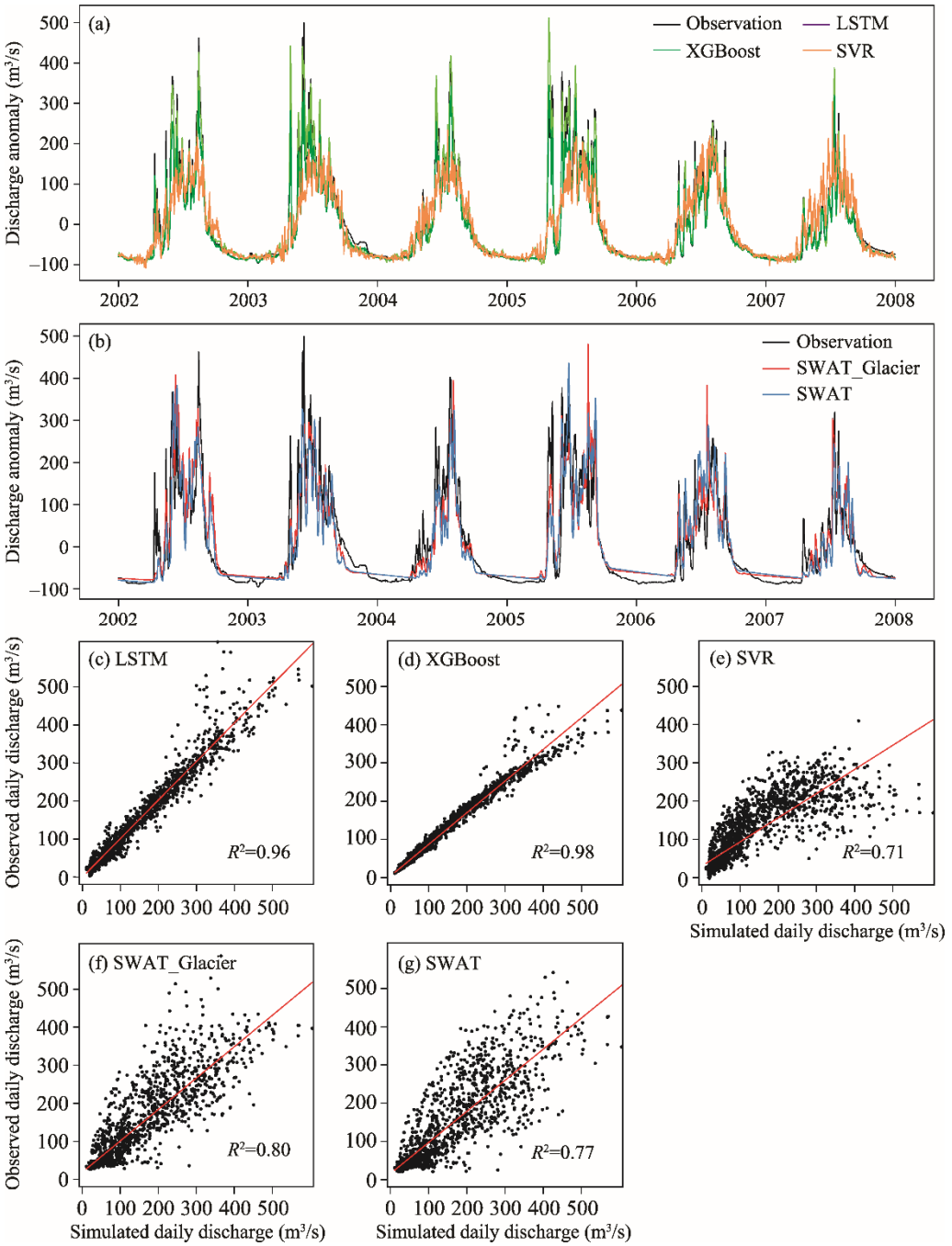

Fig. 4 Observed and simulated discharges of the five models in the Toxkan River Basin during the training period. (a), daily observations and simulations of the machine learning methods (LSTM, XGBoost, and SVR); (b), simulation performance of traditional hydrological models (SWAT and SWAT_Glacier); (c)–(g), scatterplots of observed and simulated values for each model, with R^2 representing the correlation coefficient.

Table 1 Evaluation of the LSTM, XGBoost, SVR, SWAT_Glacier and SWAT models in daily discharge simulation during the training period (2002–2007) and testing period (2008–2011) in the Kumaric and Toxkan river basins

Period	Model	Kumaric River Basin				Toxkan River Basin			
		NS	PBIAS (%)	R^2	RMSE	NS	PBIAS (%)	R^2	RMSE
Training	LSTM	0.96	–0.78	0.97	43.45	0.95	1.57	0.96	23.50
	XGBoost	0.96	–11.02	0.98	41.64	0.94	–13.17	0.98	27.23
	SVR	0.80	–5.53	0.82	89.97	0.69	–8.30	0.71	60.81
	SWAT_Glacier	0.82	0.94	0.83	85.07	0.80	–0.02	0.80	49.06
	SWAT	0.74	–10.64	0.75	103.09	0.76	–3.64	0.77	53.33
Testing	LSTM	0.90	–2.60	0.90	60.26	0.71	–0.96	0.73	51.80
	XGBoost	0.77	–3.59	0.77	86.13	0.55	–2.36	0.56	64.66
	SVR	0.79	2.41	0.79	81.80	0.55	1.53	0.56	64.62
	SWAT_Glacier	0.79	–5.40	0.80	83.06	0.50	9.90	0.58	68.76
	SWAT	0.67	–21.49	0.73	103.07	0.36	10.58	0.49	83.11

Note: LSTM, long short-term memory; XGBoost, extreme gradient boosting; SVR, support vector regression; SWAT, Soil and Water Assessment Tool; SWAT_Glacier, an extended SWAT model with glacier melting mechanism; NS, Nash-Sutcliffe efficiency coefficient; PBIAS, percentage bias; R^2, correlation coefficient; RMSE, root mean square error.

Compared with traditional hydrological models, the LSTM model not only performs calculations faster, but also is more accurate in simulation. Although SWAT_Glacier with the glacier dynamic module significantly improved in performance compared to the original SWAT model, its simulation of peak flow was still slightly lower than that of the LSTM model. Overall, these results indicated that LSTM shows major advantages compared to the other four methods and is highly applicable in simulating discharge in the Kumaric River Basin.

According to Figure 3c–e, although XGBoost was slightly inferior to LSTM, it was better than SVR. Moreover, according to the results of the evaluation indices shown in Table 1, the values of NS and R^2 of XGBoost were higher than 0.95, RMSE was 41.64, and the relative deviation of PBIAS was 11.02%. During the testing period, the NS and R^2 of XGBoost were higher than 0.75, and the PBIAS was only –3.59%, showing that all the evaluation indices were within the acceptable range. This performance was overall satisfactory, indicating that XGBoost was feasible for use in simulating discharge in the Kumaric River Basin. However, XGBoost's performance in the testing period was not as good as in the training period, demonstrating that XGBoost was prone to overfitting discharge. Still, it did show advantages in cross validation. Based on our experimentation, we found that when neglecting time series and randomly dividing the testing and training periods according to proportion, the NS value improved, remaining above 0.90 during the training periods and above 0.85 during the testing period. It is worth noting that because the division of the factor dataset is inconsistent with the other methods, it is not presented here.

Despite its earlier successes, SVR did not perform as well as LSTM or XGBoost during the training period, giving NS and R^2 of 0.80 and PBIAS of –5.53%. On the other hand, during the testing period, SVR performed much better, showing NS and R^2 of 0.79 and PBIAS of only 2.41%. In general, the overall performance of the SVR model was similar to that of the SWAT_Glacier model for the Kumaric River Basin, both of which were significantly better than the original SWAT model. As can be seen from the evaluation indices, the traditional hydrological SWAT model actually gave the worst performance among the five methods, and its performance in the training and testing periods was far lower than that of the other four methods. This clearly demonstrated that the SWAT model had dramatic limitations when applied to alpine watersheds and is thus unable to accurately describe the glacio-hydrological process.

For the Toxkan River Basin, Figure 4 presented the performance of the three mentioned machine learning algorithms and two hydrological models during the training period. It is obvious that the simulation of baseflow by the five methods was relatively consistent and that the fit was good. However, the performances of the models in the simulation of summer peaks differed greatly from each other. The LSTM model simulation results were closest to the observed values, with NS=0.95, R^2=0.96, PBIAS=1.57%, and RMSE=23.50. Further, LSTM had a high correlation coefficient and

relatively small deviation that can accurately describe the daily variations of discharge. The simulation performance of XGBoost was slightly lower than that of LSTM, and the relative deviation was relatively large. On the other hand, SVR performed poorly and the NS and R^2 values were the lowest, so it was difficult to capture any peak value trends. Compared with the SWAT model, the accuracy of the SWAT_Glacier model was improved, with NS inching up from 0.76 to

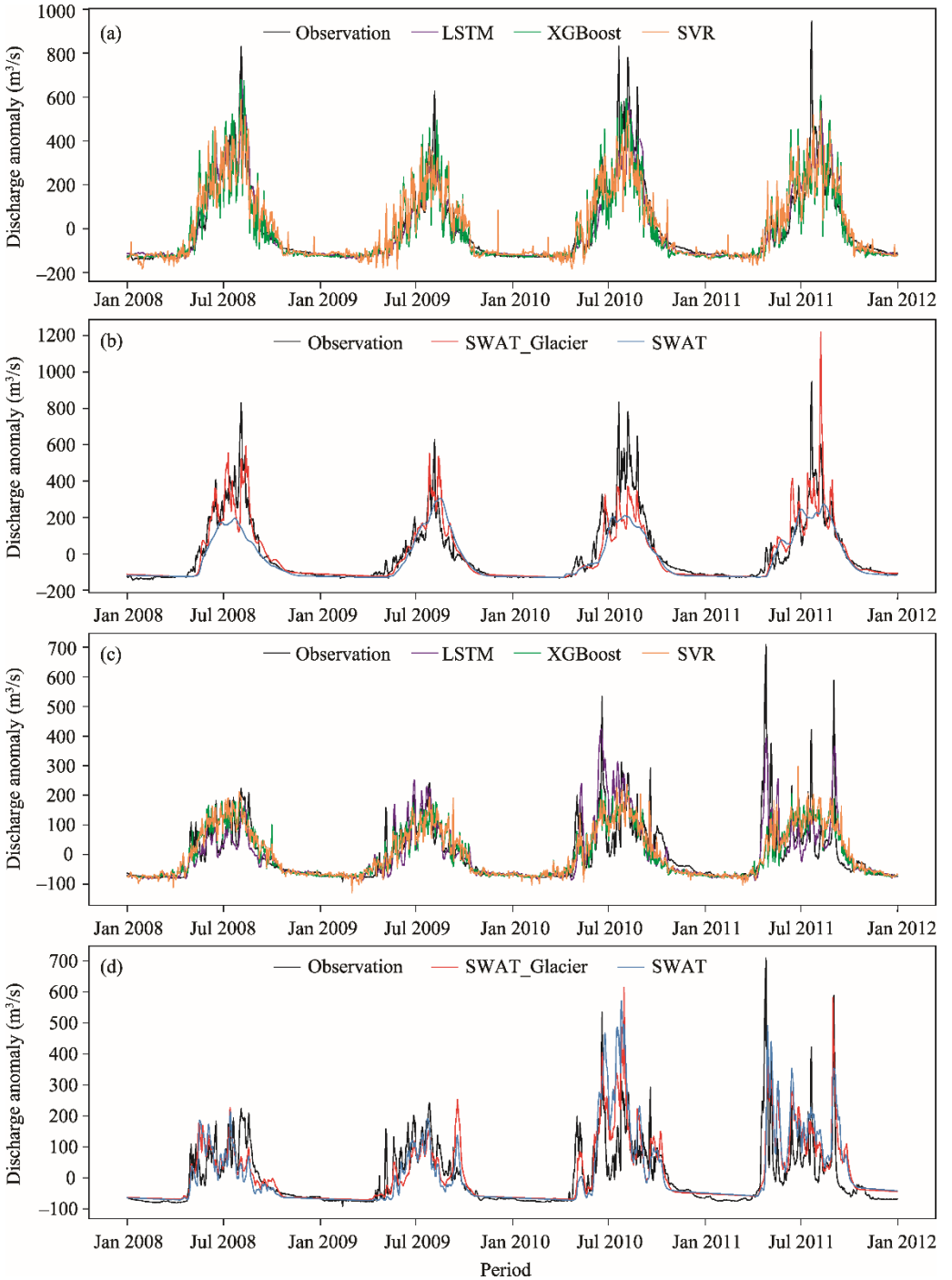

Fig. 5 Predictions of the three machine learning algorithms (LSTM, XGBoost, and SVR) and two hydrological models (SWAT_Glacier and SWAT) for discharge in the Kumaric River Basin (a and b) and Toxkan River Basin (c and d) during the testing period

0.80, PBIAS changing from –3.64% to –0.02%, and RMSE decreasing from 53.33 to 49.06. However, compared to LSTM, the performance of SWAT_Glacier was clearly inferior, which was more obvious in the testing period.

For the Toxkan River Basin, Figure 5c and d presented the forecast discharge of the five models. The LSTM model had the highest prediction accuracy, with NS=0.71, R^2=0.73, PBIAS= –0.96%, and RMSE=51.80. Both NS and R^2 values were above 0.70 in the testing period, and their performance was significantly better than that of the other models. Compared with the SWAT_Glacier model, NS was increased by 0.21, the absolute value of PBIAS was decreased by 8.94%, R^2 was increased by 0.15, and RMSE was decreased by 16.96 for the performance of LSTM model. The prediction results of LSTM were considered to be satisfactory for the model. The prediction accuracy of XGBoost and SVR was similar, giving NS of 0.55 and a generally consistent relative deviation. A similar phenomenon also occurred in the Kumaric River Basin, where the performance of the two hydrological models was not as good as that of the three classical algorithms. Once again, the SWAT model gave the worst performance, with NS and R^2 at the lowest and PBIAS and RMSE at the largest.

In order to test the robustness of various methods and analyze their performance in relation to discharge prediction on a monthly scale, we calculated the relative errors of the five models, as shown in Figure 6. The relative error was calculated by the ratio of the error of the simulation and the observation to the observed data. A positive value represents overestimation, while a negative value indicates underestimation.

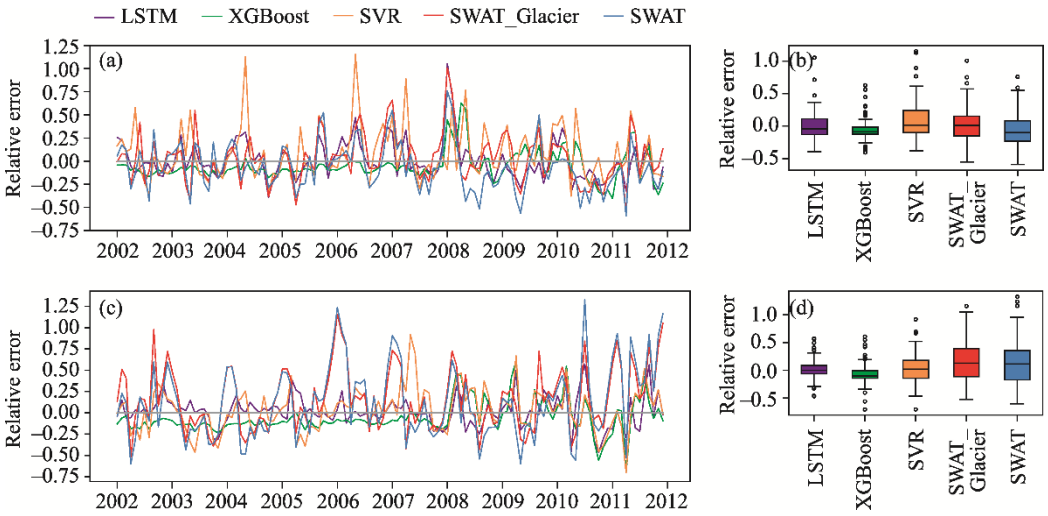

Fig. 6 Distribution of relative errors generated by machine learning algorithms (LSTM, XGBoost, and SVR) and hydrological models (SWAT_Glacier and SWAT) on a monthly scale in the Kumaric River Basin (a and b) and Toxkan River Basin (c and d). The box and whisker plots show the five-number summary of a set of data: the minimum score, first (lower) quartile, median, third (upper) quartile, and the maximum score. The center represents the middle 50%, or 50th percentile of the data set, and is derived using the lower and upper quartile values. The median value is displayed inside the "box". The maximum and minimum values are displayed with vertical lines ("whiskers") connecting the points to the center box. The circles represent the outliers. The box and whisker plots have the same meaning as Figure 8.

Figure 6a and c indicated that the relative errors of the five models were mainly distributed in the range from –0.25 to 0.25. Specifically, the LSTM error was quite small during the entire training and testing periods. XGBoost underestimated discharge, with a small error in the training period but a large error during the testing period. This reveals that XGBoost is prone to overfitting discharge, resulting in a poor fitting effect for the testing period. Unlike LSTM or XGBoost, SVR was more likely to overestimate and thus showed a larger error in monthly discharge simulation.

Compared with the other three algorithms, the SWAT_Glacier and SWAT models had larger errors. Figure 6b and d showed that the analysis results of the Toxkan River Basin were similar to

those of the Kumaric River Basin, and that the simulations between the two hydrological models were larger. Compared with the other models, LSTM showed a smaller error, which was more clearly demonstrated in the boxplot in Figure 6. It can be seen that although the median of the five models was near the baseline, the relative errors of the LSTM model for the two basins were mainly distributed in a fairly small range, with 0.0 as the center. Although the boxplot of XGBoost was a narrow box representing the middle 50% of the data, the overall distribution was underestimated. For both SWAT_Glacier and SWAT, in addition to the large box width, the difference between the maximum value and the minimum value was large. This indicates that the relative error between the simulated and observed values was large.

In general, the discharge simulations of the five models for the Kumaric River Basin were better than those for the Toxkan River Basin. The LSTM method in particular showed a great potential to simulate and forecast discharge in both basins. Although XGBoost had a good performance in the training period, its application was limited by the issue of overfitting, and the performance of SVR was not as good as that of LSTM. Compared with the SWAT model, the SWAT_Glacier model exhibited a better simulation ability, but the performance of the SWAT_Glacier model was still much lower than that of the LSTM model. Overall, the SWAT model gave the poorest performance among the five models, revealing its unsuitability for discharge simulations in glaciated alpine regions.

5 Discussion

In comparing the evaluation indices, we found that the five models performed better in the Kumaric River Basin than in the Toxkan River Basin. One of the key reasons for the differences in performance is the unevenness of the distribution of meteorological stations in the Toxkan River Basin. Because the basin is located in a remote mountain area characterized by undulating hills and valleys, it is challenging to install meteorological stations, and the maintenance cost is prohibitive. Hence, there are only two meteorological stations in the basin, covering an area of 1.92×10^4 km^2, and these stations are located on the border of the basin.

The elevation range of the Toxkan River Basin is 1884–5934 m, with an average altitude of 3558 m (Fig. 7). One of the meteorological stations, called Akqi, is located in the downstream plains near the outlet of the basin, at an altitude of 1985 m. Its location allows it only a weak representation of the climate condition in the upstream area. Therefore, its contribution to the simulation is very minor. The other meteorological station, called Tuergate, is located outside the basin but at a relatively high altitude of 3504 m. Due to the mentioned complex topographical features of the basin, the data from the Akqi and Tuergate stations cannot accurately reflect the basin's overall spatiotemporal variations of temperature and precipitation. This is a critical failure, as temperature and precipitation are considered the most crucial factors affecting the process of glacier and snow accumulation and melting (Chen et al., 2017). Previous studies have shown that temperature plays a dominant role in the discharge of the Kumaric River Basin (Ji et al., 2019), whereas in the Toxkan River Basin, the discharge is affected by both temperature and precipitation (Duethmann et al., 2015).

Figure 8 showed the distribution of precipitation, maximum temperature, and minimum temperature for meteorological observation stations in the two basins. The Aksu, Koilu, and Tianshan stations are located in the Kumaric River Basin, with elevations of 1103, 2800, and 3614 m, respectively. Figure 8c and e presented temperature variations in the three meteorological observation stations across different altitude zones. As can be seen from the figure, the changes in the maximum and minimum temperatures at the three stations were relatively consistent, and the variation range was small. This indicated the high quality of the observed temperature data, and since temperature played a key role in the discharge of the Kumaric River Basin, all five models had good performances.

The Akqi and Tuergate meteorological stations are located in the Toxkan River Basin. Figure 8d and f presented changes in the maximum and minimum temperatures, respectively, with overall temperature variations being relatively minor. However, as shown in Figure 8b, precipitation at the two stations varied greatly from June to October. The uncertainty in the precipitation data caused

challenges in discharge simulations in the Toxkan River Basin, which is further impacted by the combined effects of temperature and precipitation. The observation data from the two stations reveal some defects in their reflection of the precipitation distribution for the basins. The imprecision of the temperature and precipitation data can lead to significant errors when altitude is changed, limiting the applicability and performance of the hydrological models. Therefore, neither the traditional hydrological models nor the deep learning methods are effective.

In addition, the variations in discharge in glaciated alpine watersheds are not only closely related to changes in temperature, precipitation, and other important climate factors, but also affected by changes in glacier and snow distribution, glacier dynamics, groundwater, and other external factors. Therefore, integrating physical models with deep learning methods for further exploration might be a promising future research direction.

The simulation of peak flow is an insurmountable problem in hydrological models, and it is almost impossible for such simulations to be as good as baseflow. However, LSTM has obtained highly satisfactory results for both baseflow and peak flow, potentially making this method ideal for the study of regional floods. Especially in glacier-dominated basins with floods caused by glacier lake outbursts, the LSTM approach is expected to deepen our understanding of floods in general, which is an important direction for future research. The accurate prediction of floods can provide important decision-making evidence for reducing flood-related losses.

As demonstrated in this study, the LSTM method provides a new perspective in regional hydrological research, especially in glaciated alpine regions where meteorological data are scarce. LSTM sheds new light on discharge simulations through the enhancement of result accuracy compared to traditional hydrological models that are dependent on a large number of physical parameters. Furthermore, data-driven methods do not necessarily require prior knowledge of the watershed hydrological system, as proposed by Razavi and Coulibaly (2013).

It should be admitted nonetheless that all these methods have their own advantages and disadvantages. The model-dependent methods can interpret the physical mechanisms of the hydrological process by calibrating hydrological parameters with physical significance in a specific hydrological model. However, the operation of these types of models relies not only on large amounts of spatial attribute data (such as land use and spatial distribution of soil), but also on high-quality meteorological observation data, resulting in complicated time- and cost-consuming computational methods.

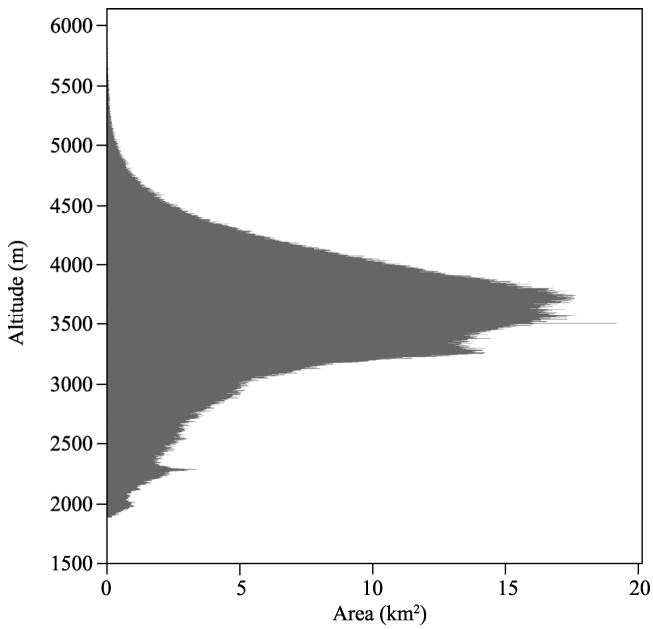

Fig. 7 Relationship between basin area and altitude in the Toxkan River Basin

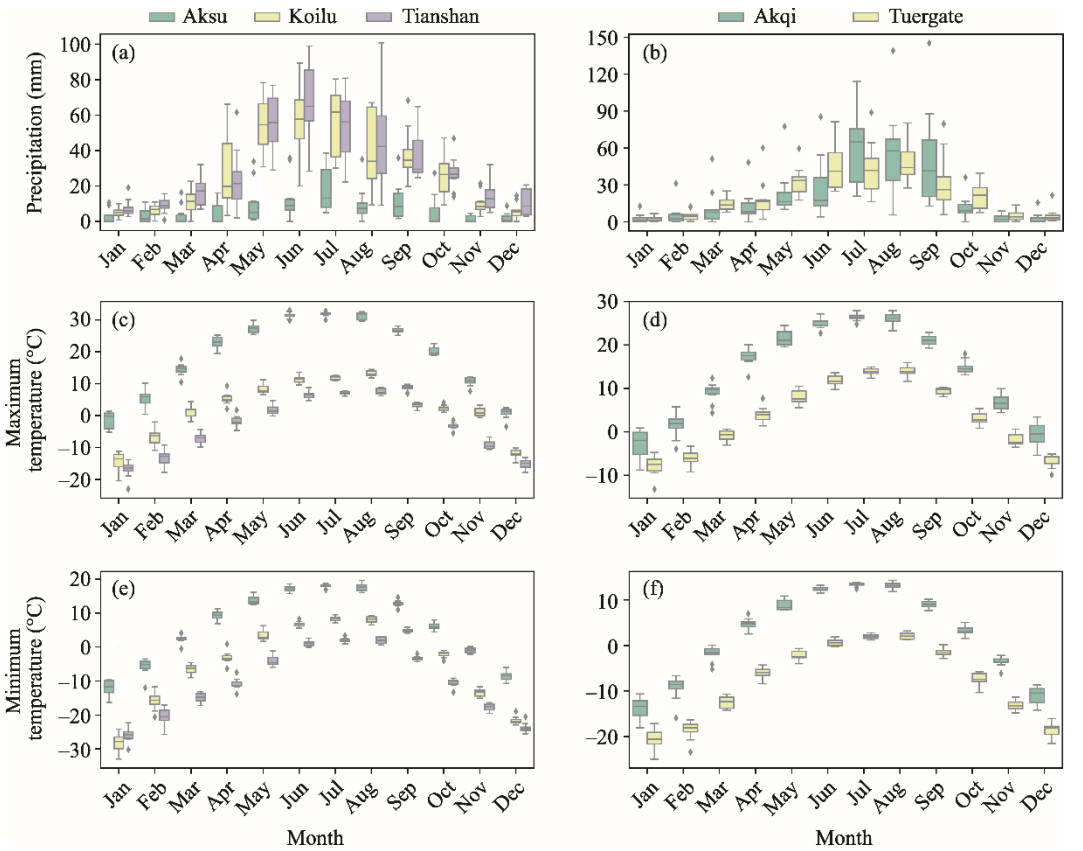

Fig. 8 Boxplot of precipitation, maximum temperature, and minimum temperature at meteorological observation stations in the Kumaric River Basin (a, c, and e) and Toxkan River Basin (b, d, and f)

In this regard, the LSTM method has substantial advantages. For alpine watersheds where the discharge is mainly dominated by meteorological factors, only meteorological forcing data are needed as input factors in the neural network, and the underlying surface spatial attribute data with little contribution to discharge can be ignored. LSTM also does not introduce complex hydrological parameters, which avoids the issue of "different parameters with the same effect" that is usually encountered in hydrological models. Furthermore, the LSTM algorithm does not focus on the physical mechanisms of the hydrological cycle and thus ignores the causal relationship between the variables in the study area. Compared with the distributed hydrological models, the LSTM method, which only relies on observation variables with time series and learns the mapping relationship between discharge and input characteristics, uses less input variables to obtain better simulation results. This technique, with its unprecedented accurate prediction capabilities, provides a new strategy for hydrological simulations of data-sparse alpine river basins.

6 Conclusions

Three machine learning methods (LSTM, XGBoost, and SVR) and two physical hydrological models (SWAT_Glacier and SWAT) for simulating discharge in snow- and glacier melt-dominated basins in alpine regions (the Tianshan Mountains) were presented and tested in this study. Our comparison of the five approaches demonstrated that the artificial neural network algorithm LSTM has incomparable advantages over the others. The LSTM model has a strong adaptability for simulating discharge in data-sparse glaciated-dominated basins. The data-driven algorithm improves the accuracy of discharge simulation by learning the mapping relationship between discharge and meteorological data. The other two machine learning methods (XGBoost and SVR), however, show certain limitations, e.g., XGBoost tends to overfit and SVR has poor simulation

abilities. The performance of the SWAT_Glacier model is inferior to that of the LSTM model, but it has significantly improved from the original SWAT model. In our study, simulations from the SWAT_Glacier model showed a high consistency with observations, while the SWAT model without the benefit of the glacier dynamic module had difficulty in accurately describing the glacio-hydrological process in glaciated alpine basins. Compared with hydrological models, the LSTM algorithm has a higher simulation accuracy with simple input data and low calculation costs. Therefore, the application of LSTM to hydrological research in data-sparse alpine basins marks a major step forward.

In the future, we might be considering combining machine learning algorithms with hydrological models to establish the physics-aware machine learning methods, which can systematically analyze the physical mechanisms in glaciated alpine basins. This is an interesting and meaningful direction.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (U1903208, 41630859, 42071046). The authors wish to express great thanks to Prof. YANG Jing from National Institute of Water and Atmospheric Research in New Zealand for his guidance on hydrological models. We are grateful to the anonymous reviewers and editors for their insightful and constructive comments that helped improve the manuscript.

References

Aggarwal S K, Goel A, Singh V P. 2012. Stage and discharge forecasting by SVM and ANN techniques. Water Resources Management, 26: 3705–3724.

Ajami N, Gupta H, Wagener T, et al. 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. Journal of Hydrology, 298: 112–135.

Arnold J G, Srinivasan R, Muttiah R S, et al. 1998. Large area hydrologic modeling and assessment—Part 1: Model development. JAWRA Journal of the American Water Resources Association, 34: 73–89.

Bengio Y, Simard P, Frasconi P. 1994. Learning Long-Term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5: 157–166.

Chen T Q, Guestrin C. 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: Special Interest Group on Management of Data, 785–794.

Chen X, Long D, Hong Y, et al. 2017. Improved modeling of snow and glacier melting by a progressive two-stage calibration strategy with GRACE and multisource data: How snow and glacier meltwater contributes to the runoff of the Upper Brahmaputra River basin? Water Resources Research, 53: 2431–2466.

Chen Y N, Takeuchi K, Xu C C, et al. 2006. Regional climate change and its effects on river runoff in the Tarim Basin, China. Hydrological Processes, 20: 2207–2216.

Chen Y N, Li W H, Deng H J, et al. 2016. Changes in Central Asia's water tower: past, present and future. Scientific Reports, 6: 35458, doi: 10.1038/srep35458.

Chen Y N, Li W H, Fang G H, et al. 2017. Review article: Hydrological modeling in glacierized catchments of Central Asia – status and challenges. Hydrology and Earth System Sciences, 21: 669–684.

Cheng C T, Feng Z K, Niu W J, et al. 2015. Heuristic methods for reservoir monthly inflow forecasting: A case study of Xinfengjiang Reservoir in Pearl River, China. Water, 7: 4477–4495.

Duethmann D, Bolch T, Farinotti D, et al. 2015. Attribution of streamflow trends in snow and glacier melt-dominated catchments of the Tarim River, Central Asia. Water Resources Research, 51: 4727–4750.

Duethmann D, Menz C, Jiang T, et al. 2016. Projections for headwater catchments of the Tarim River reveal glacier retreat and decreasing surface water availability but uncertainties are large. Environmental Research Letters, 11: 054024, doi: 10.1088/1748-9326/11/5/054024.

Fan Y T, Chen Y N, Li W H. 2014. Increasing precipitation and baseflow in Aksu River since the 1950s. Quaternary International, 336: 26–34.

Fang K, Shen C P, Kifer D, et al. 2017. Prolongation of SMAP to spatiotemporally seamless coverage of continental U.S. using a deep learning neural network. Geophysical Research Letters, 44, doi: 10.1002/2017gl075619.

Fang K, Shen C P. 2019. Near-real-time forecast of satellite-based soil moisture using long short-term memory with an adaptive data integration kernel. Journal of Hydrometeorology, 21: 399–413.

Feng D P, Fang K, Shen C P. 2020. Enhancing streamflow forecast and extracting insights using Long-Short Term Memory networks with data integration at continental scales. Water Resources Research, 56: e2019WR026793, doi: 10.1029/2019WR026793.

Finger D, Vis M, Huss M, et al. 2015. The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments. Water Resources Research, 51: 1939–1958.

Fu M L, Fan T C, Ding Z A, et al. 2020. Deep learning data-intelligence model based on adjusted forecasting window scale: Application in daily streamflow simulation. IEEE Access, 8: 32632–32651.

Gao S, Huang Y F, Zhang S, et al. 2020. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology, 589: 125188, doi: 10.1016/j.jhydrol.2020.125188.

Greff K, Srivastava R K, Koutník J, et al. 2017. LSTM: A search space odyssey. IEEE Transactions on Neural Networks and Learning Systems, 28: 2222–2232.

Greve P, Burek P, Wada Y. 2020. Using the Budyko framework for calibrating a global hydrological model. Water Resources Research, 56: e2019WR026280, doi: 10.1029/2019WR026280.

Hochreiter S, Schmidhuber J. 1997. Long short-term memory. Neural Computation, 9: 1735–1780.

Hsu K, Gupta H V, Sorooshia S. 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resources Research, 31: 2517–2530.

Hu C H, Wu Q, Li H, et al. 2018. Deep learning with a Long Short-Term Memory networks approach for rainfall-runoff simulation. Water, 10: 1543, doi: 10.3390/w10111543.

Hu Y C, Yan L, Hang T, et al. 2020. Stream-flow forecasting of small rivers based on LSTM. arXiv e-prints. arXiv:2001.05681. https://arxiv.org/abs/2001.05681.

Immerzeel WW, van Beek L, Bierkens M. 2010. Climate change will affect the Asian water towers. Science, 328: 1382–1385.

Immerzeel W W, Lutz A F, Andrade M, et al. 2020. Importance and vulnerability of the world's water towers. Nature, 577: 364–369.

Ji H P, Fang G H, Yang J, et al. 2019. Multi-objective calibration of a distributed hydrological model in a highly glacierized watershed in Central Asia. Water, 11: 554, doi: 10.3390/w11030554.

Jodar J, Carpintero E, Martos-Rosillo S, et al. 2018. Combination of lumped hydrological and remote-sensing models to evaluate water resources in a semi-arid high altitude ungauged watershed of Sierra Nevada (Southern Spain). Science of the Total Environment, 625: 285–300.

Kingma D, Ba J. 2015. Adam: A Method for Stochastic Optimization. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego: Computational and Biological Learning Society.

Kratzert F, Klotz D, Brenner C, et al. 2018. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrology and Earth System Sciences, 22: 6005–6022.

Kratzert F, Klotz D, Herrnegger M, et al. 2019a. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. Water Resources Research, 55: 11344–11354.

Kratzert F, Klotz D, Shalev G, et al. 2019b. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences, 23: 5089–5110.

Kratzert F, Klotz D, Hochreiter S, et al. 2020. A note on leveraging synergy in multiple meteorological datasets with deep learning for rainfall-runoff modeling. Hydrology and Earth System Sciences Discussion. [Preprint]. https://doi.org/10.5194/hess-2020-221, in review.

Li W, Kiaghadi A, Dawson C. 2021. High temporal resolution rainfall–runoff modeling using long-short-term-memory (LSTM) networks. Neural Computing and Applications, 33: 1261–1278.

Li Z L, Shao Q X, Xu Z X, et al. 2010. Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia watershed in northwest China. Journal of Hydrology, 385: 76–83.

Luo X G, Yuan X H, Zhu S, et al. 2019. A hybrid support vector regression framework for streamflow forecast. Journal of Hydrology, 568: 184–193.

Maier H R, Jain A, Dandy G C, et al. 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environmental Modelling & Software, 25: 891–909.

Michaud J, Sorooshian S. 1994. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. Water Resources Research, 30: 593–605.

Ni L L, Wang D, Wu J F, et al. 2020. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. Journal of Hydrology, 586: 124901, doi: 10.1016/j.jhydrol.2020.124901.

Qin J, Yang K, Liang S L, et al. 2009. The altitudinal dependence of recent rapid warming over the Tibetan Plateau. Climatic

Change, 97: 321, doi: 10.1007/s10584-009-9733-9.

Razavi T, Coulibaly P. 2013. Streamflow prediction in ungauged basins: Review of regionalization methods. Journal of Hydrologic Engineering, 18: 958–975.

Reichstein M, Camps-Valls G, Stevens B, et al. 2019. Deep learning and process understanding for data-driven earth system science. Nature, 566: 195–204.

Rudin C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1: 206–215.

Shen C P. 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research, 54: 8558–8593.

Shen C P, Laloy E, Elshorbagy A, et al. 2018. HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. Hydrology and Earth System Sciences, 22: 5639–5656.

Shen Y J, Shen Y J, Fink M, et al. 2018a. Unraveling the hydrology of the glacierized Kaidu Basin by integrating multisource data in the Tianshan Mountains, Northwestern China. Water Resources Research, 54: 557–580.

Shen Y J, Shen Y J, Fink M, et al. 2018b. Trends and variability in streamflow and snowmelt runoff timing in the southern Tianshan Mountains. Journal of Hydrology, 557: 173–181.

Tarasova L, Knoche M, Dietrich J, et al. 2016. Effects of input discretization, model complexity, and calibration strategy on model performance in a data-scarce glacierized catchment in Central Asia. Water Resources Research, 52: 4674–4699.

Viviroli D, Kummu M, Meybeck M, et al. 2020. Increasing dependence of lowland populations on mountain water resources. Nature Sustainability, 3: 917–928.

Xiang Z R, Yan J, Demir I. 2020. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resources Research, 56, doi: 10.1029/2019wr025326.

Yang T, Sun F B, Gentine P, et al. 2019. Evaluation and machine learning improvement of global hydrological model-based flood simulations. Environmental Research Letters, 14: 114027, doi: 10.1088/1748-9326/ab4d5e.

Yang T T, Asanjan A A, Welles E, et al. 2017. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. Water Resources Research, 53: 2786–2812.

Zheng Z S, Ma Q, Jin S C, et al. 2019. Canopy and terrain interactions affecting snowpack spatial patterns in the Sierra Nevada of California. Water Resources Research, 55: 8721–8739.